# CHARACTERIZATION OF MYCOSIN FAMILY PROTEASES, NOVEL DRUG TARGETS OF *MYCOBACTERIUM TUBERCULOSIS*

A Senior Scholars Thesis

by

Hilary Jean Baird

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for designation as

UNDERGRADUATE RESEARCH SCHOLAR

April 2007

Major: Biochemistry

# CHARACTERIZATION OF MYCOSIN FAMILY PROTEASES, NOVEL DRUG TARGETS OF *MYCOBACTERIUM TUBERCULOSIS*

A Senior Scholars Thesis

by

Hilary Jean Baird

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor: James C. Sacchettini
Associate Dean for Undergraduate Research: Robert C. Webb

April 2007

Major: Biochemistry

# ABSTRACT

Characterization of Mycosin family proteases, Novel Drug Targets of *Mycobacterium tuberculosis* (April 2007)

Hilary J. Baird
Department of Biochemistry
Texas A&M University

Research Advisor: Dr. James C. Sacchettini
Department of Biochemistry

Tuberculosis is the world's leading cause of death by infectious disease. Antibiotic resistance and HIV co-infection is increasing at an alarming rate. Mycosins-1-5 are subtilisin-like serine proteases within the periplasmic space of the tuberculosis bacterial cell. However, to date, Mycosin-1 is thought to be the most interesting because it is only expressed after TB infection and is thought to be essential to its virulence. The role of Mycosins-2-4 is not yet known. These factors make Mycosin-1 a novel drug target. This study aims to clone and characterize Mycosin-1 for further investigation as a drug target. The mycosin-1 and mycosin-2 genes were successfully cloned for later use in expression studies. Mycosin-2 has been included in the cloning process because of the conservation of the Mycosins' active sites. The most potent inhibitor will be able to bind all five Mycosins. Several sequence alignments have also been included to help characterize the Mycosin-1 protein. The evidence suggests that Mycosin-1 is a typical subtilisin-like protease, allowing the characterization of the protein. A homology model

has been built to gain further insight into the protein and for later use in virtual inhibitor screening.

# DEDICATION

This thesis is dedicated to my best friend and better half Jessica Noyes. I know I would have made you proud.

# ACKNOWLEDGMENTS

I would like to thank Nilofar Mohaideen for providing technical assistance throughout the project. Also, Anup Aggarwal for so much help, advice, and editing. He gave me immeasurable encouragement when the going got tough. Stephanie Swanson for editing and answering a million questions, as well as all others who offered their advise and support.

I would also like to thank Dr. James Sacchettini and the Sacchettini laboratory for supporting my efforts in the Undergraduate Research Scholars Program and for providing resources and funding.

# NOMENCLATURE

| | |
|---|---|
| *Mtb* | *Mycobacterium tuberculosis* |
| WHO | World Health Organization |
| MDR-TB | Multi-Drug Resistant Tuberculosis |
| XDR-TB | Extreme-Drug Resistant Tuberculosis |
| Snm | Secretion in Mycobacteria |
| MycP1 | Mycosin-1 |
| MycP2 | Mycosin-2 |
| DNA | Deoxyribonucleic acid |
| PCR | Polymerase Chain Reaction |
| LB Media | Luria-Bertani Media |
| PDB | Protein Data bank |

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER I

# INTRODUCTION

*Mycobacterium tuberculosis* is the single leading cause of death by infectious disease in the world. The WHO (World Health Organization) reported that in 2004, 1.7 million deaths were a result of *Mtb* infections (1). Although the rate of infection is decreasing in developed areas of the world, it is increasing by 0.6% per year on a global scale (1). The infection is spread through the air and individuals with active, untreated, infections will on average infect 10-15 people per year. It is also estimated that only half of *Mtb* cases were reported by health-care systems in 2004 and just 82% of these finished treatment (2). *Mtb* has the ability to survive for long periods in the host mandating extended drug treatments and thus resulting in low compliance. A recent article also claimed that "one-third of the world's population is infected with TB, and hundreds of thousands of children will become TB orphans this year" (3).

*Mtb* is reaching an alarming magnitude in underdeveloped regions of the world such as Africa. As the number of people with AIDS increase, so does the rate of *Mtb*. In fact, it is the leading cause of death of individuals who have AIDS. Also, AIDS patients are 50 times more likely to develop *Mtb* than those who are HIV negative. Approximately one third of people with HIV are co-infected with *Mtb* and 90% of those will die within months of the *Mtb* infection without treatment (4).

---

This thesis follows the style of *Journal of Biological Chemistry*.

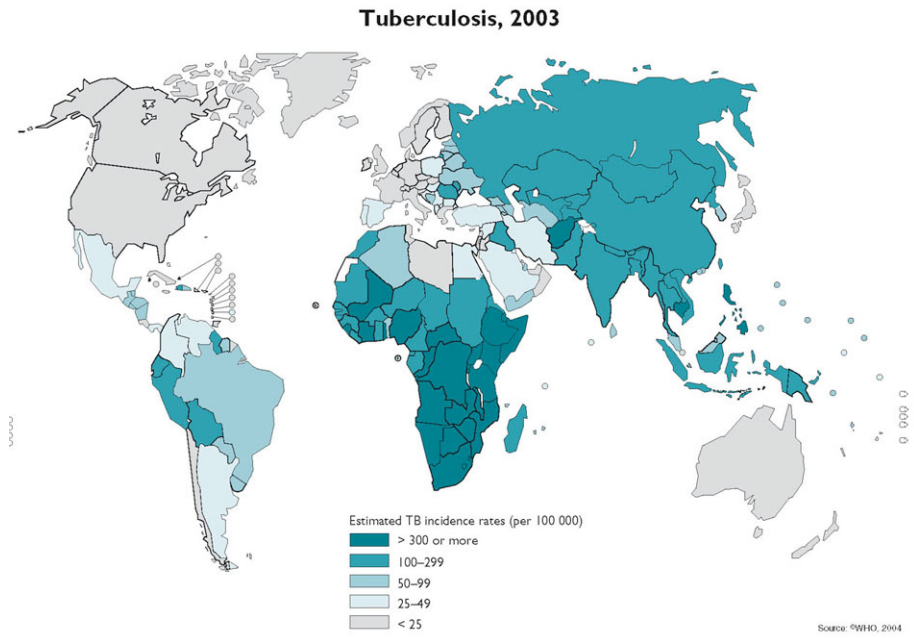


Fig. 1. **Incidences of reported Tuberculosis in 2003.** (1)

## Drug Resistance

In the past years, strains of drug resistant *Mtb* have become prevalent. In fact, resistance is so wide spread that it is now being classified as multi-drug resistant (MDR-TB) and extreme-drug resistant (XDR-TB). Indeed, strains of *Mtb* have even evolved resistance to all major anti-TB drugs available. The infection can still be cured with extensive chemotherapy, however a need for a more readily available treatment is evident. Little is known about the actual mechanism that controls the interactions between the host and *Mtb* and even less is known on how such interactions affect the life cycle of the host or the pathogen. Therefore, novel targets for new drug treatment are needed.

**Mycosin-1**

The mycosins are a family of 5 genes present in *Mtb*. They are subtilisin-like serine proteases with a highly conserved catalytic triad (Asp, His, Ser) (5). They each have a C-terminal transmembrane region and an N-terminal signal peptide, represented in figure 2. Mycosin-1 (MycP1), Rv3883c, is an extra cellular protein that is membrane and cell wall associated. It is most likely subject to cleavage of the signal peptide region following secretion from the cell (5). MycP1 is expressed after infection of *Mtb*, making it a potential drug target.

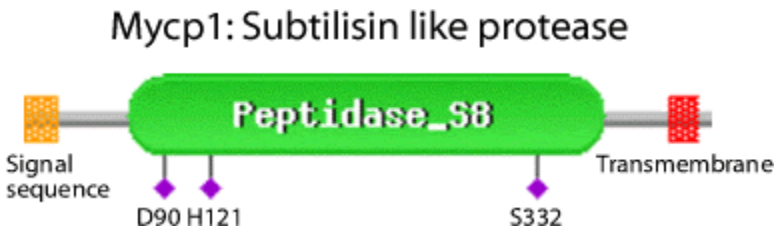


Fig. 2. **A representation of the MycP1 protein**. This figure denotes the active site residues, the signal sequence and the transmembrane region.

The Mycosin proteins are highly conserved, shown by the sequence alignment in figure 3. The active site residues are denoted using orange stars. Most importantly, the active sites of the proteins are conserved. This is significant because an inhibitor molecule that binds to one Mycosin will have a higher probability to inhibit all of them, allowing all five genes to be targeted for drug design.

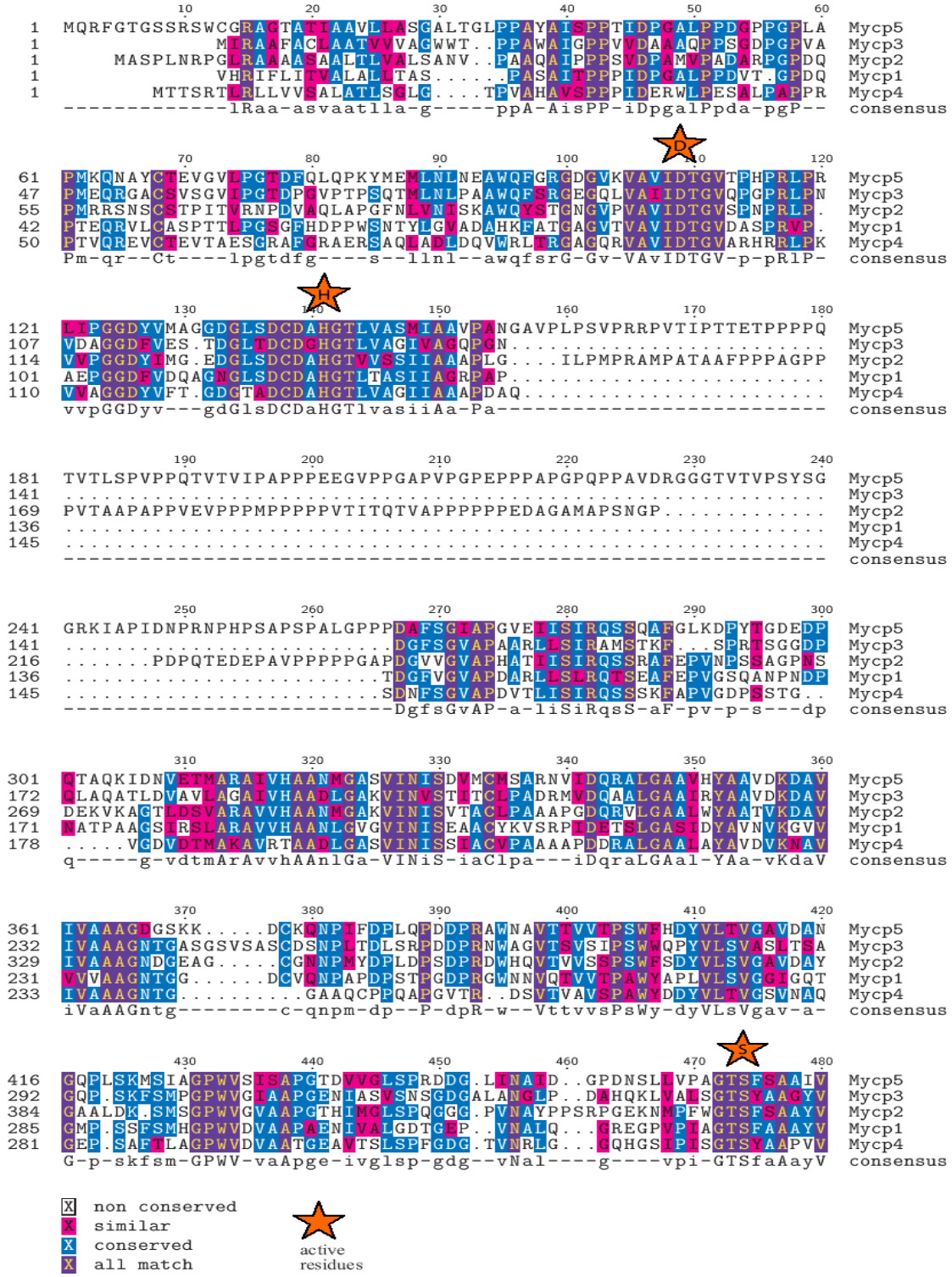


Fig. 3. **Sequence alignment of the five mycosin proteins.** Sequence produced using Clustalw, version 3.2 and Textshade (6) (7).

In Mtb, there are eight regions of difference (termed RD1-8) between the genome of the virulent and non-virulent strains (8). RD1 is known to contain the *esxA*/*esxB* operon which encodes two predominant Mtb antigens, ESAT-6 and CFP-10. Several of the genes surrounding the *esxA*/*esxB* operon were deleted from *Mycobacterium bovis* on the path to obtaining the attenuated *M. bovis* BCG vaccine strain (8). An alternate protein secretion system, called the Snm (secretion in mycobacteria) pathway, is encoded in Mtb. This system includes RD1 as well as several surrounding genes that are conserved to the region. The proteins encoded by these surrounding genes are thought to aid in the secretion of the ESAT-6 and AFP-10 antigens (8). The most notable of these is Mycosin-1 (MycP1), because it was found to be required for the functioning of the Snm pathway. Because of MycP1's location in the periplasm, it is hypothesized that MycP1 acts as a regulator of Snm secretion, either breaking down an Snm inhibitor or activating an Snm protein (8).

# CHAPTER II

# METHODS

## Truncation and Primer Design

*Mycosin-1*

MycP1 has an N-terminal signal sequence and C-terminal transmembrane region as shown in figure 3. Previous attempts to express the full-length protein yielded insoluble protein products (Jim Sacchettini, Personal communication). Generally, signal peptides and transmembrane helices consist of high proportions of hydrophobic residues. Removing both these hydrophobic tail regions should greatly enhance the solubility of the protein. The protein was therefore truncated at the $49^{th}$ and the $417^{th}$ amino acid residue to increase the solubility after expression. The active site residues occur at amino acids 101 (Asp), 122 (His), and 336 (Ser); leaving a sufficient number of amino acids before the active site residues, making this a safe truncation.

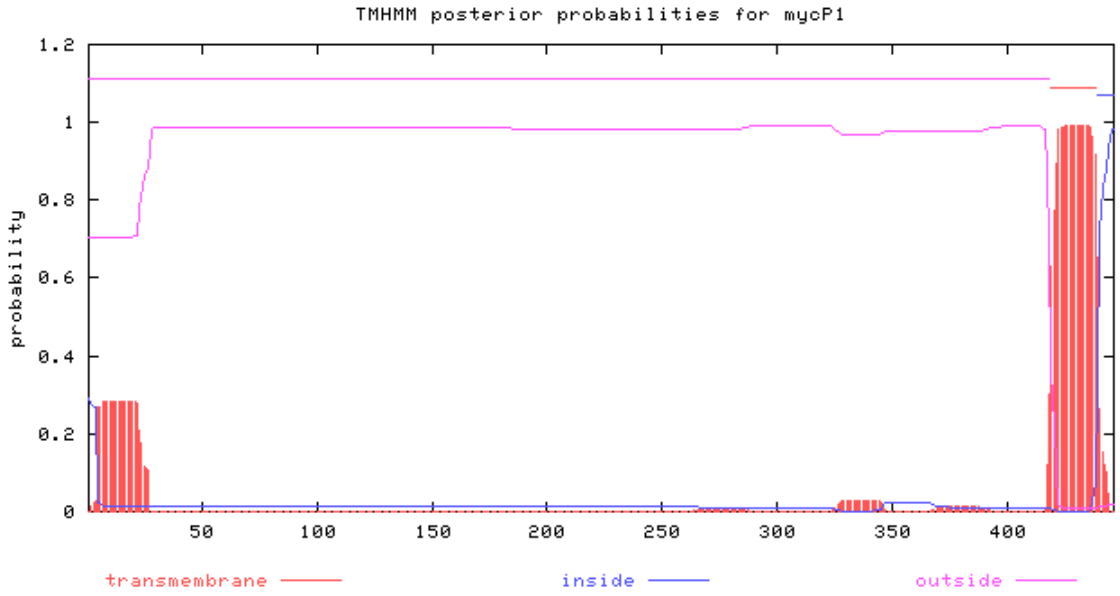


Fig. 4. **The prediction of transmembrane helices for the mycP1 amino acid sequence.** The TMHMM server was used. The sequence was truncated at amino acid residues 49 and 417 (9).

The *E. coli* vector pET-28b ((Novagen, EMD Biosciences)) was used with the restriction enzymes Nhe1 and HindIII. Two reverse primers were designed; one without a stop codon and the other with a stop codon inserted at the end of the truncated mycP1 sequence. This allowed an N-terminal His-tag and both N- and C-terminal His-tags to be inserted onto the protein sequence. The primer sequences used were:

Forward (using Nhe1): 5'AGCCTAGCTAGCGATCAGCCTACCGAACAGCG 3'

Reverse, no stop (using HindIII):

5'AGCCTAAAGCTTGCGACGATCGGGACCCGGCTC 3'

Reverse, stop (using HindIII):

5'AGCCTAAAGCTTTCAGCGACGATCGGGACCCGGCTC 3'

*Mycosin-2*

Due to the conservation among the mycosins, cloning of mycosin-2 (mycP2), Rv3886c, was also tried. This protein also has hydrophobic rich N-terminal signal sequence and C-terminal transmembrane region, as seen in figure 4. The protein was truncated at the 56$^{th}$ and 507$^{th}$ amino acids thus removing both the hydrophobic regions.

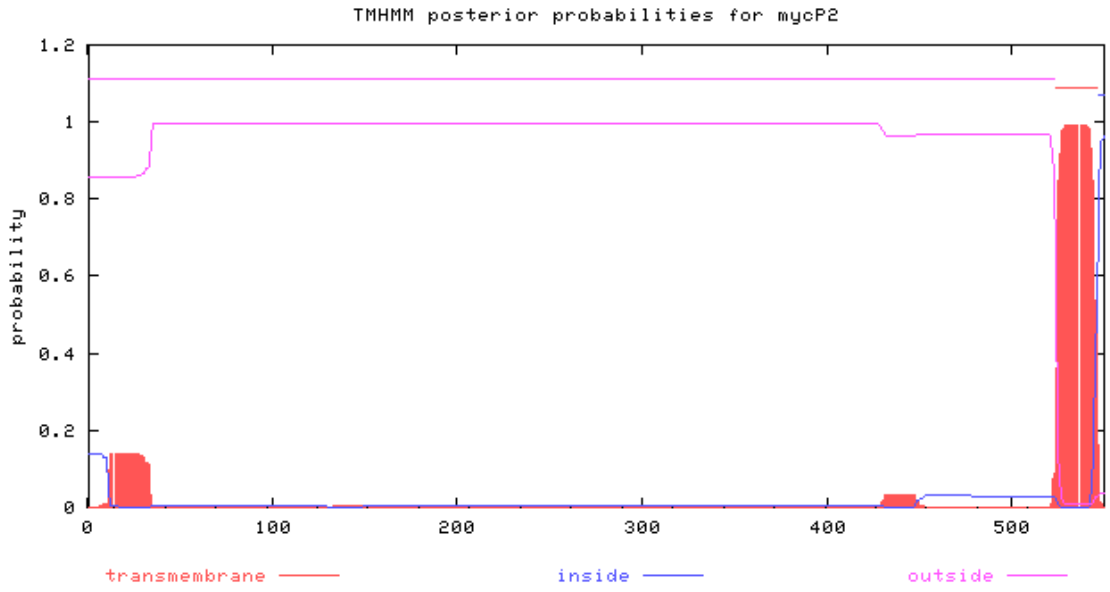


Fig. 5. **The prediction of transmembrane helices for the mycP2 amino acid sequence.** The protein was truncated at amino acids 56 and 507. The active site residues occur at amino acids 103 (Asp), 133(His), and 435(Ser). The small hydrophobic region occurring between amino acids 425 and 450 could not be truncated without interfering with the active site of the protein. However, it should not interfere with expression and the truncation made is considered safe (9).

The primers for mycP2 were designed to allow proteins with N-terminal, C-terminal, and both N- and C-terminal His-tags to be produced. The *E. coli* vector pET-28b (Novagen, EMD Biosciences) was used with the restriction enzymes Nde1 and HindIII to produce protein sequences with the N-terminal and the N- and C-terminal His-tags.

The *E. coli* vector pET-30b (Novagen, EMB Biosciences) was used to produce a protein sequence with only C-terminal His- tag. The primers used were:

Forward (using Nde1): 5'AGCCTACATATGGCGCCGCTCCAACAGCTGCTCCA 3'

Reverse no stop (using HindIII):

5'AGCCTAAACATTCCGGCTCTGTGCACCCGGGGC 3'

Reverse stop (using HindIII):

5'AGCCTAAAGATTTCACCGGCTCTGTGCACCCGGGGC 3'

**Cloning**

The mycP1 and mycP2 gene fragments were amplified using PCR; a FailSafe PCR kit (Epicentre Biotechnologies) was used. An agarose DNA gel was run to check the PCR products. The amplified PCR products were purified using a PCR purification kit (Qiagen). The respective vector and PCR products were double digested using the restriction enzymes previously defined. The outcomes of the double digestions were run on an agarose DNA gel to ensure the correct cutting by the restriction enzymes. The mycosin DNA fragments were then ligated with the respective linearized *E. coli* vector and transformed into Novablue *E. coli* competent cells. A Rapid DNA Ligation Kit (Roche) was used for the ligation.

The Novablue cells were plated onto LB media containing the antibiotic kanamycin (50 μg/ml) and allowed to incubate overnight at 37°C. A single colony from the plate was transferred into liquid LB media, containing kanamycin, and grown at 37°C for several

hours. A plasmid miniprep kit (Qiagen) was used to extract the *E. coli* vectors containing the cloned mycosin gene. The mycP1 and mycP2 plasmids were then sent for DNA sequencing to ensure the correct inserts.

# CHAPTER III

# RESULTS

## Cloning

The DNA encoding hydrophobic regions of mycP1 and mycP2 were truncated to promote soluble expression of the proteins in *E. coli* cells. After truncation, mycP1 was approximately 1100 base pairs and mycP2 was approximately 1500 base pairs. MycP1 and mycP2 genes were successfully cloned using *Mtb* genomic DNA (H37Rv strain). The primer sequences for N-terminal His-tags were used for the PCR amplification. The PCR products were run on a DNA agarose gel to ensure the best buffer conditions for producing DNA fragment were used, as seen in figure 5. Although products were seen in conditions D, G, and K, only conditions D and G were used for mycP1 and mycP2 respectively. The DNA concentrations of these samples were estimated to be 82.5ng/µl for mycP1 and 57.5ng/µl for mycP2.

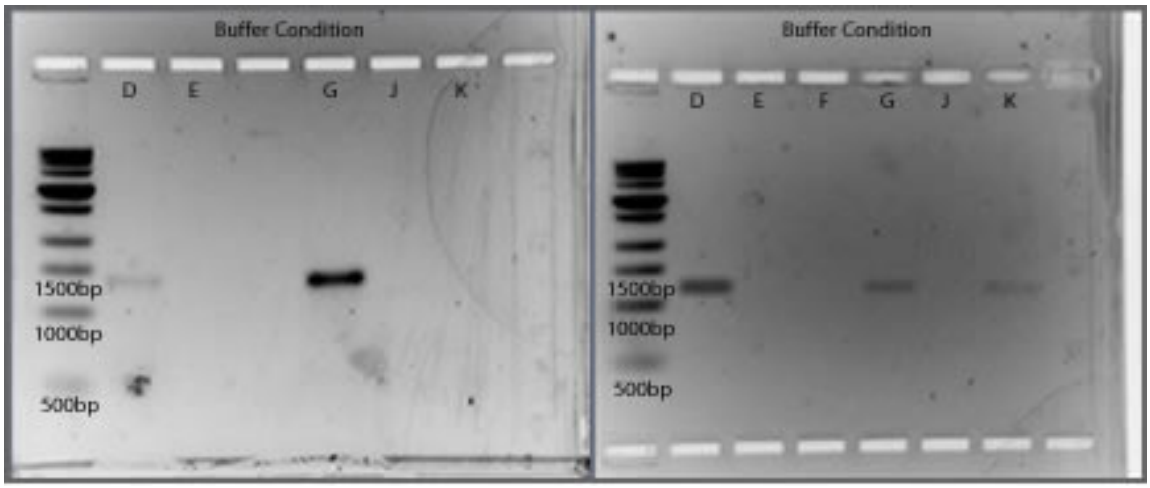


Fig. 6. **PCR products of mycP1 and mycP2.** Left: PCR products of mycP2, the darkest band corresponds to buffer condition G. Right: PCR products of mycP1, the darkest band corresponds to buffer condition D.

However, several additional PCR reactions were run to produce more samples. A Phusion High-fidelity PCR kit (New England Bio Labs) was used with two separate reaction mixes to optimize the insert amplification produced. The agarose gel of the PCR products can be seen in figure 6. Reaction mix one, using the Failsafe buffer condition G, only produces a significant amount of mycP2 DNA fragments. The reaction mix two, using 5x fusion buffer and 10mM dNTP's, did not produce a significant amount of either gene. A test DNA (1300 bp) was used to ensure the PCR reaction worked.

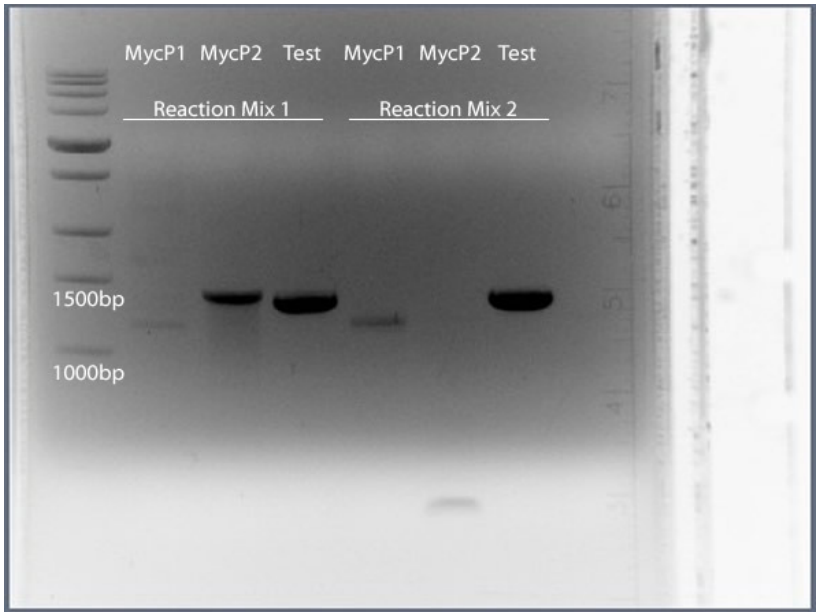


Fig. 7. **PCR products using Phusion High-fidelity PCR kit.** A test was used both reaction mixes to ensure the efficiency of the PCR. The mycP1 gene was not significantly produced in either mix, while mix one produced a significant amount of mycP2. The low band in the mycP2, reaction mix two lane is most likely the primers used in the reaction.

The pET-28b vector and DNA fragments were double digested with restriction enzymes. DNA fragments from two PCR's were used for each gene to ensure the effectiveness of the double digestion. The NEB buffer 2 and BSA were used for maximum efficiency. The section cut out of the pET-28b vector was too large, greater than 100bp, to be removed using the PCR purification kit. Therefore, the linearized vector had to be purified by running it onto an agarose gel and then gel purifying the vector band by excising it from the gel. The linearized vector was approximately 5200 bp and estimated to be 95 ng/µl.

The two samples of mycP1 and mycP2 genes were initially ligated with the linearized vector using a 1:3 (vector to insert) ratio. However, after transforming the ligation products into Novablue (a non-expression strain of *E. coli*), no colonies were produced on the LB plates. This could have been due to a failed ligation or failed transformation. The ligation reactions were repeated using the ratios 1:3 as well as 1:4 (vector to insert). The transformation into Novablue cells was successful with both ratios. The colonies produced using the ratio 1:3 were chosen for further experiment.

The Novablue cells were grown in LB media and harvested to extract the plasmids containing the mycP1 and mycP2 genes. The extracted plasmids were sent for DNA sequencing as well as tested for the presence of insert. A double digestion using the original restriction enzymes was run for each of the plasmids. The agarose gel can be seen in figure 8. Samples corresponding to the two separate PCR are shown. The double digestion of the mycP2 sample shown in lane three revealed that the ligation reaction failed. The plasmid must have recircularized without the mycP2 gene insert. This is proven by the absence of any addition bands in the lane as well as the size of the digested plasmid. From the figure 8 it is clearly evident that the double digestion was successful as the vector backbone bands are all of same expected size.

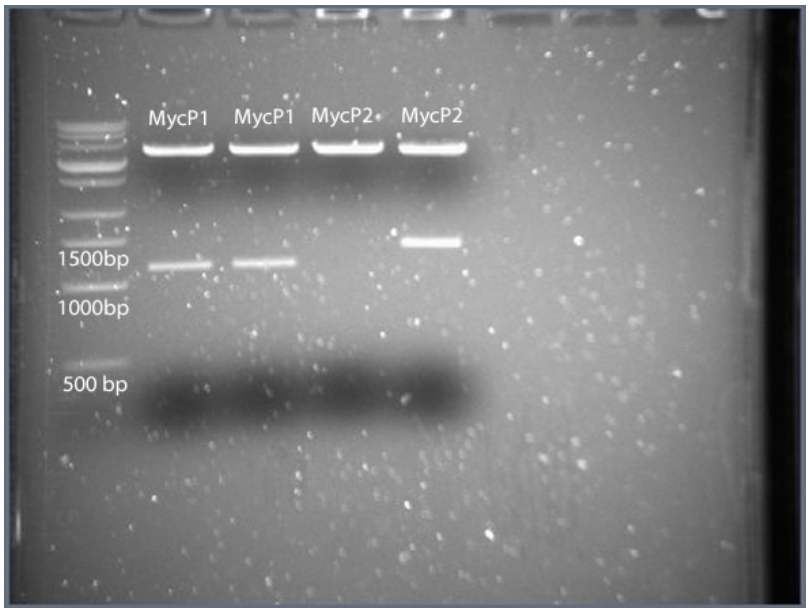


Fig. 8. **Double digestion verification of the gene inserts.**

The gene inserts were also verified using PCR. The reaction mix one was used because only it produced visible bands for both of the genes. Instead of using the genomic DNA of *Mtb* as a template, the plasmids extracted from Novablue cells were used. The results of the PCR verification are shown in figure 9. The failure of the mycP2 gene to insert into the vector is further proven. No products can be seen from that reaction. This is most likely due to the absence of the gene insert in the plasmid used as the template.

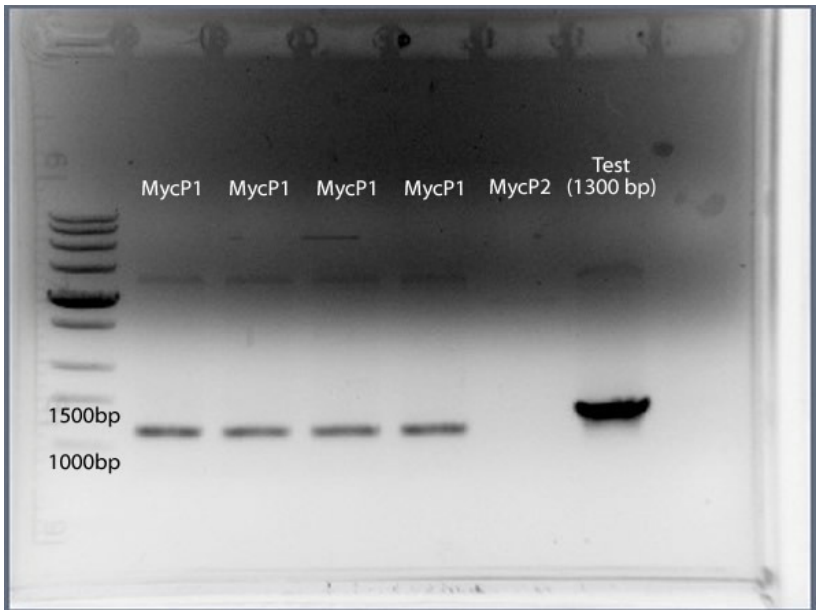


Fig. 9. **PCR verification of the gene inserts.**

## Mycosin Homologs

MycP1 is a subtilisin-like serine protease that is thought to be essential to the life cycle of *Mtb*. Subtilisins can be found in viruses, bacteria, as well as eukaryotes. For this reason, the MycP1 protein sequence was BLASTed against the non-redundant protein sequence database to identify homologs in other species. The results showed that the majority of homologs are contained in bacteria. Although similar sequences were found mainly in *Mycobacterium*, some were identified in *Nocardia*, *Bacillus*, and *Corynebacterium* as well as others. Table 1 identifies a sampling of the species found to contain homologs to mycP1.

| Species | PDB/Reference Number | Characteristics | Identity |
|---|---|---|---|
| Mycobacterium leprae | NP_301157.1 | Probable secreted protease | 354/446 (79%) |
| Mycobacterium smegmatis | YP_884499 | Membrane-anchored mycosin mycp1 | 321/443 (72%) |
| Nocardia farcinica | YP_117037 | Putative protease | 198/436 (45%) |
| Saccharopolyspora erythraea | YP_001108883 | Putative serine protease | 165/419 (39%) |
| Corynebacterium glutamicum | BAB97968 | Hypothetical protein | 138/390 (35%) |
| Streptomyces avermitilis | NP_823716 | Serine protease | 118/362 (32%) |
| Bacillus sp. | ZP_01168843 | Alkaline serine protease, subtilase | 104/346 (30%) |
| Erythrobacter litoralis | YP_458816 | Predicted subtilase | 108/328 (32%) |
| Lyngbya sp. | ZP_01619435 | Peptidase S8 and S53, subtilisin, kexin | 104/342 (30%) |
| Bacillus sp. | 1DBI | Thermostable Serine protease | 90/268 (33%) |

Table 1. **A sampling of the species found to contain homologs of the mycP1 protein.** The search produced several hits in mycobacterium species and only few of them are shown here for clarity (10).

A phylogenetic tree, figure 10, was constructed to examine the evolutionary distance between the bacterial homologs. The tree was constructed using an implicit alignment between the database sequences. This is based on the alignment of the sequences with the MycP1 protein.

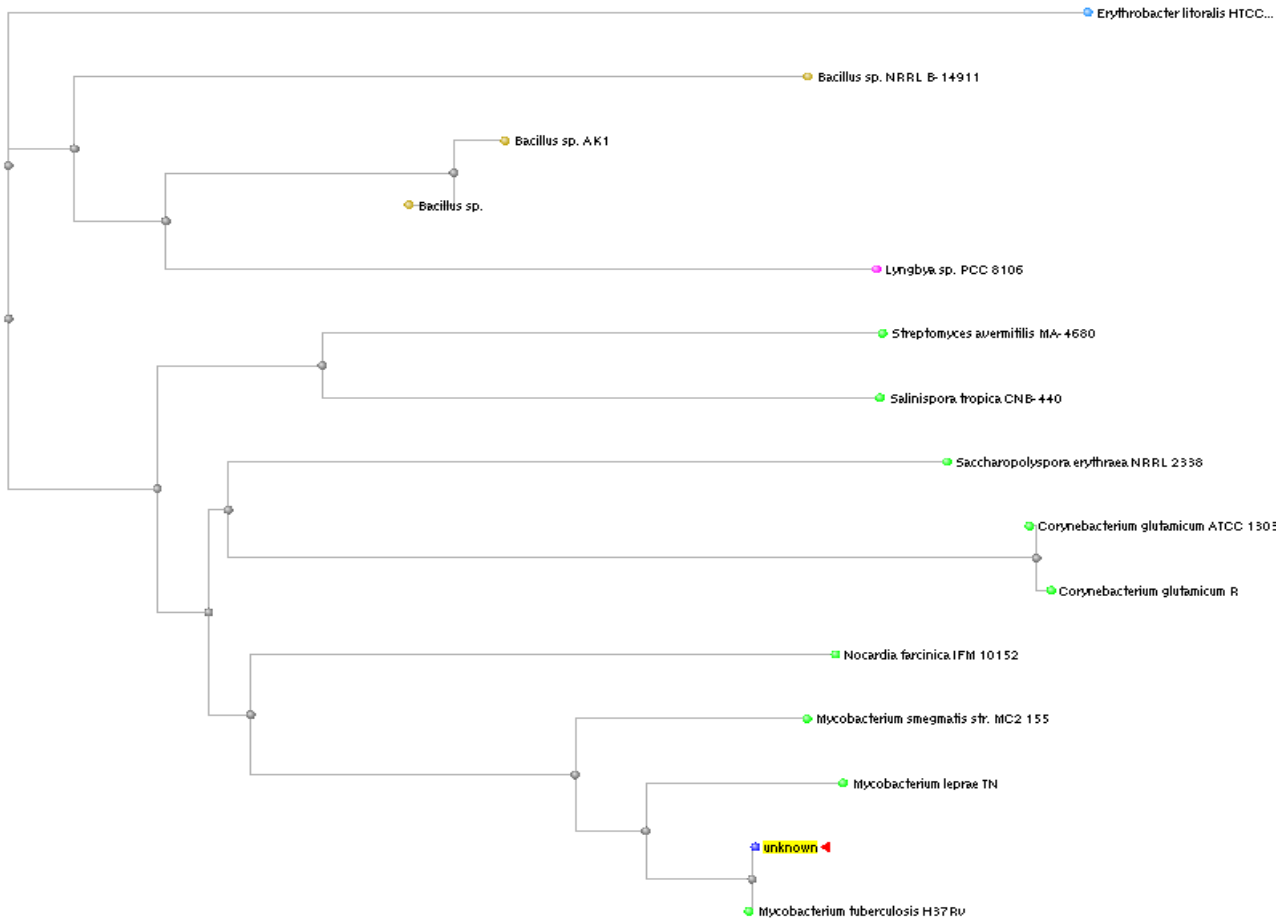


Fig. 10. **Phylogenetic tree showing the divergence of mycP1 homologs.** The protein labeled unknown in yellow highlighting is the submitted mycP1 sequence. The closest match the program made was, in fact, the mycP1 protein in *Mtb*, seen directly next to the submitted sequence. Key: high GC Gram+: green, eubacteria: brown, a-proteobacteria: blue, cyanobacteria: magenta. The tree was produced using BLAST distance tree results and a pairwise alignment (10).

The MycP1 protein was also found to have a homolog in *Homo sapiens*. The conservation in species ranging from bacteria to humans suggests that MycP1 may play an important role in the life cycle. Although the identity between human and *Mtb* proteins is low, this result is significant for drug design. It is essential that an inhibitor designed to treat an infection in humans does not interact with, and therefore, potentially

inactivate human proteins. Even though the infection may be successfully treated, the inhibitor could also cause sever and unexpected effects in a person.

The similarity of the two proteins can make inhibitor design for the MycP1 protein difficult because of the potential binding to the human homolog. It is possible, however, to specifically design the inhibitor only for the *Mycobacterium* species because the conservation between the two proteins is low. Figure 11 shows the sequence alignment of the MycP1 protein and the human homolog. It can be seen that the active site of the MycP1 protein (denoted by orange stars) is not conserved in the human homolog, even though it is also a subtilisin/kexin-like protease. This will enable the designing of an inhibitor that will only fit the MycP1 active site.

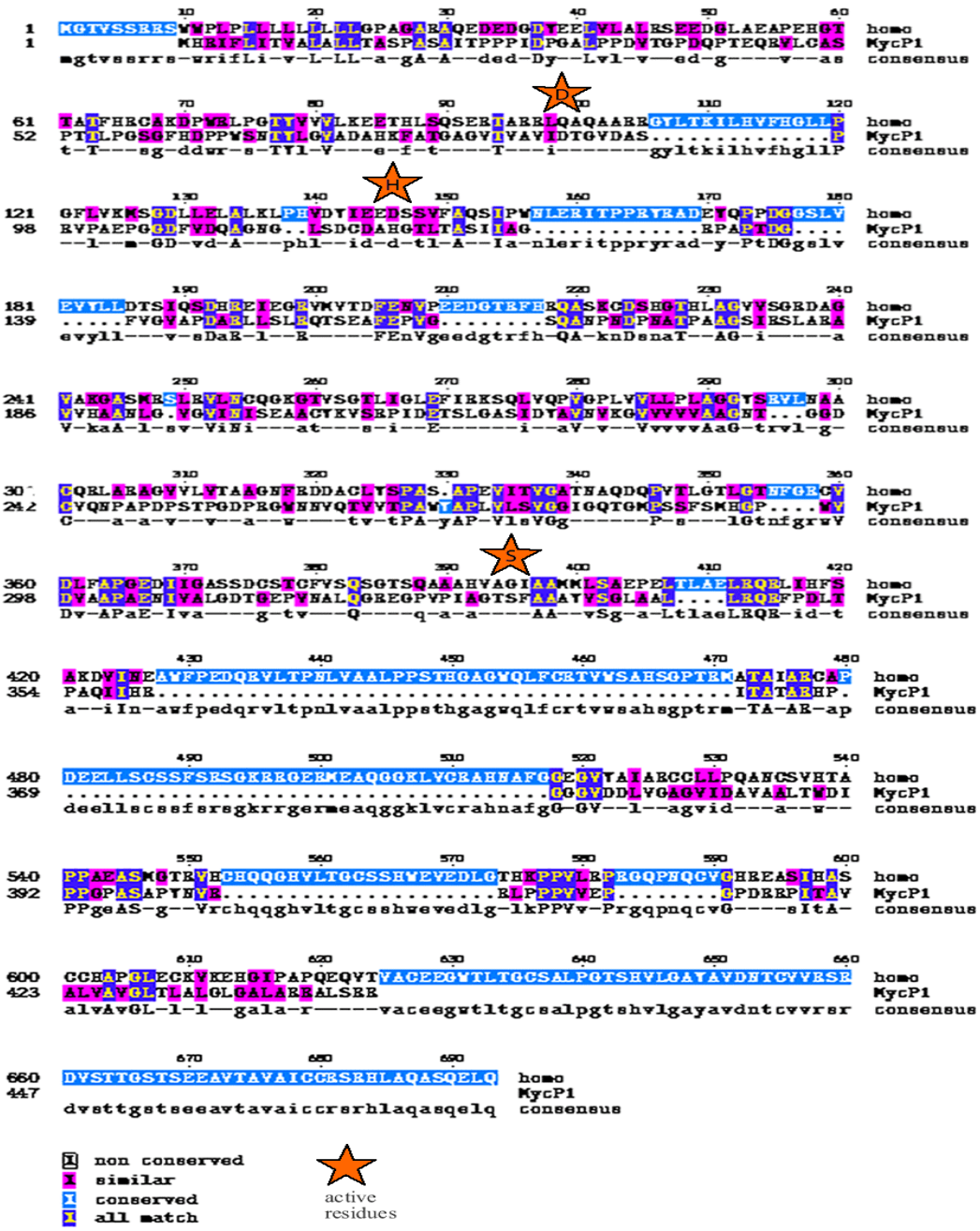


Fig. 11. **The sequence alignment of MycP1 and a human homolog.** Orange stars denote the active site of the MycP1 protein. The active site is not conserved even though the human homolog is a subtilisin/kexin-like protease as well (6) (7).

*Subtilisins*

Subtilisins are proteolytic enzymes that were originally identified in *Bacillus subtilis*, belonging to the peptidase S8 family. It is the largest class of serine proteases so far identified. A catalytic triad containing the amino acids serine, aspartic acid and histadine characterizes them. A three-element fingerprint involving the motifs surrounding the active residues identifies these proteins. Table 2 shows the fingerprint results of MycP1, subtilisin being the top and only significant match.

| **Fingerprint** | **# of Motifs** | **IdScore** | **PfScore** | **Pvalue** | **Sequence** | **Len** | **Low** | **Pos** | **high** |
|---|---|---|---|---|---|---|---|---|---|
| SUBTILISIN | 1 of 3 | 38.89 | 341 | 5.11e-06 | GAGVTVAVI DTGVDASPRVP | 20 | 23 | 473 | 81 |
| SUBTILISIN | 2 of 3 | 43.83 | 269 | 2.64e-03 | DCDAHGTLT ASIIA | 14 | 58 | 552 | 117 |
| SUBTILISIN | 3 of 3 | 49.11 | 382 | 8.84e-07 | AGTSFAAAYV SGLAALL | 17 | 212 | 864 | 329 |

Table 2. **Fingerprint results of the MycP1 protein sequence.** The motifs involved include: 1- the "region encoded by PROSITE pattern SUBTILASE_ASP (PS00136), which contains the active Asp; motif 2 spans the region encoded by PROSITE pattern SUBTILASE_HIS (PS00137), which contains the active His; and motif 3 includes the region encoded by PROSITE pattern SUBTILASE_SER (PS00138), which contains the active Ser." (11) (12)

The MycP1 homologs identified are serine proteases with most of them belonging to the subtilisin family. The conservation of motifs throughout these proteins was identified and can be seen in figure 12. The motifs within the MycP1 protein are highly conserved along the other subtilisin proteases. For this reason, it can be inferred that MycP1 will follow the general characteristics of the subtilisin family.

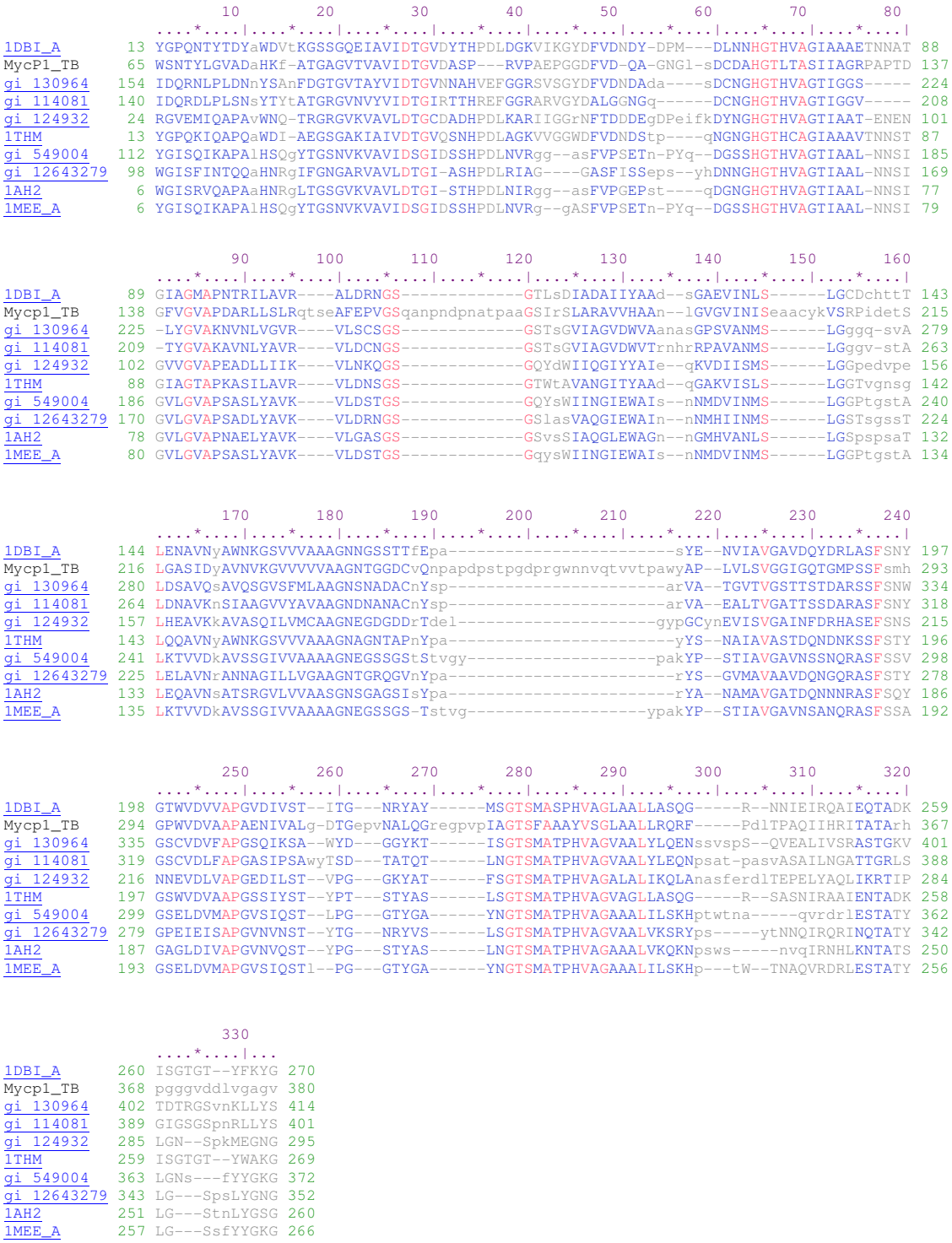


Fig. 12. **The motif alignment of MycP1 with other subtilisin proteins.** The sequence alignment is produced by the CDD server (13).

## Secondary Structure Prediction of MycP1

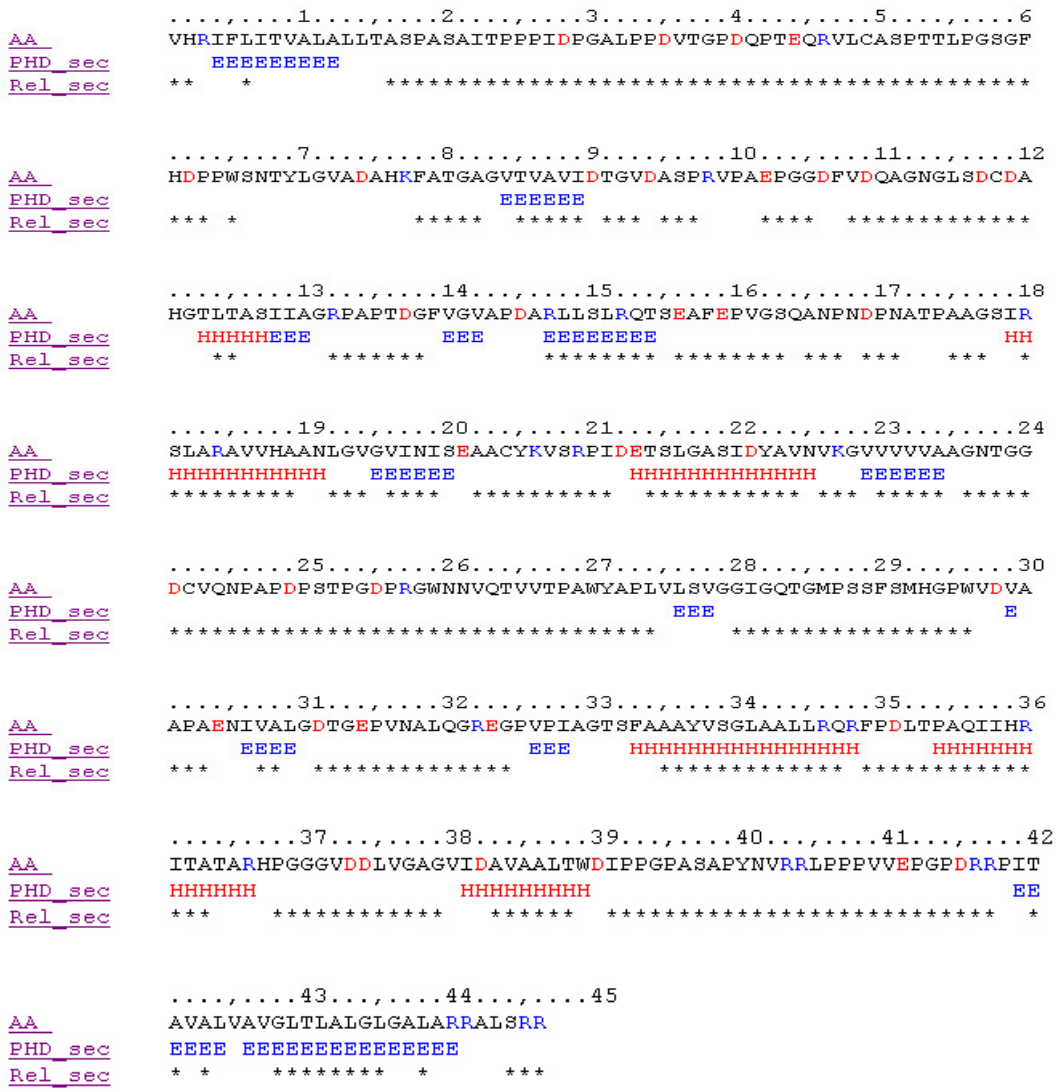


Fig. 13. **Secondary Structure prediction of MycP1.** This prediction was created using the PredictProtein server. The results shown are PHD predictions (14).

The predicted secondary structure, figure 13, of MycP1 shows an approximately even mix of beta-sheets and helixes. However, the majority of the structure is shown to be loops. This is consistent with the homology model presented in the following section.

**Homology Modeling**

Homology modeling is the process of predicting the three dimensional structure of a protein. This is accomplished by identifying a known crystal structure of a similar protein. Protein structures are more conserved than protein sequences, making this a significant tool in characterizing proteins. However, the quality of the model will depend on the quality of the sequence alignment and template structure. *Bacillus* Ak.1 protease (PDB ID 1DBI) was identified as the most accurate template for MycP1 with 41% homology (110/268 AA.). There are other more similar proteins to the MycP1 protein, although there are no known structures for the use as a template. The alignment of the two proteins can be seen in figure 14 and most importantly, the active site is conserved. The human homolog identified in figure 11 does not have a conserved active site with the bacterial mycosin proteins. This is significant because it will allow an inhibitor to be designed that will only fit into the bacterial mycosin active site, essentially not inhibiting the human homolog.

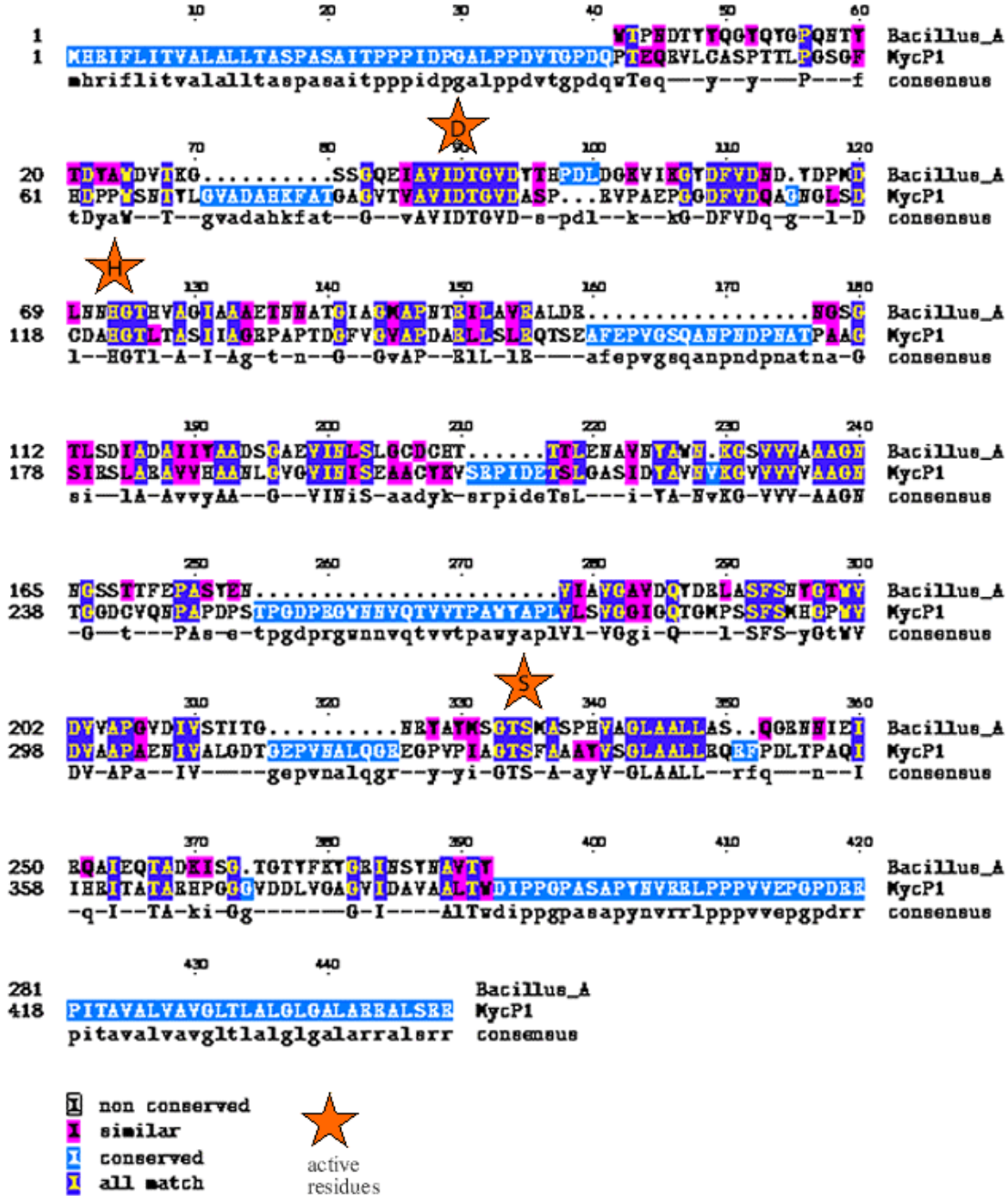


Fig. 14. **The sequence alignment of *Bacillus* Ak.1 and MycP1**. Orange stars denote the active site residues (6) (7).

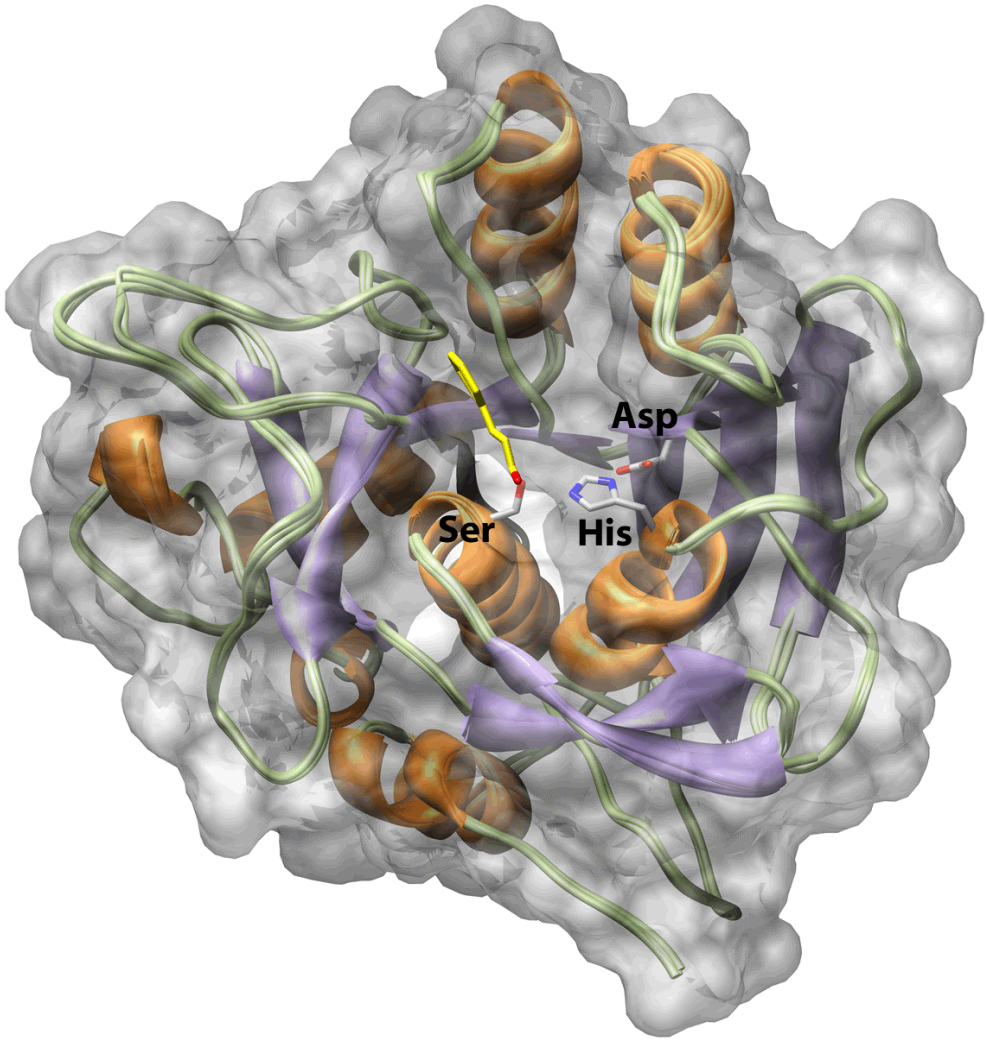


Fig. 15. **Homology model of mycP1 generated using Swiss Molder.** (15)

The homology model was built using the Swiss homology model sever and the MODELLER6.0 program and is seen in figure 15. This model shows the active site of the putative MycP1 protein with an inhibitor docked in it. Applying the information known about the *Bacillus* Ak.1 protease and subtilisins in general to MycP1, the active site can be characterized. Most subtilisins contain the catalytic triad on the C-terminal end of a core β-sheet (16). The active site cleft is most likely located along the surface

of the molecule. Here, two loops generally join two strand/helix pairs in subtilisins (16). In all known crystal structures of subtilisin proteins, calcium-binding sites have been identified and are thought to contribute to the stability of the proteins. However, the number of ions and specificity varies between families of subtilisins. The *Bacillus* Ak.1 protease contains three calcium ions and one sodium ion. The last binding site containing the sodium ion is thought to be highly interchangeable with a calcium ion in the *Bacillus* Ak.1 protease (16).

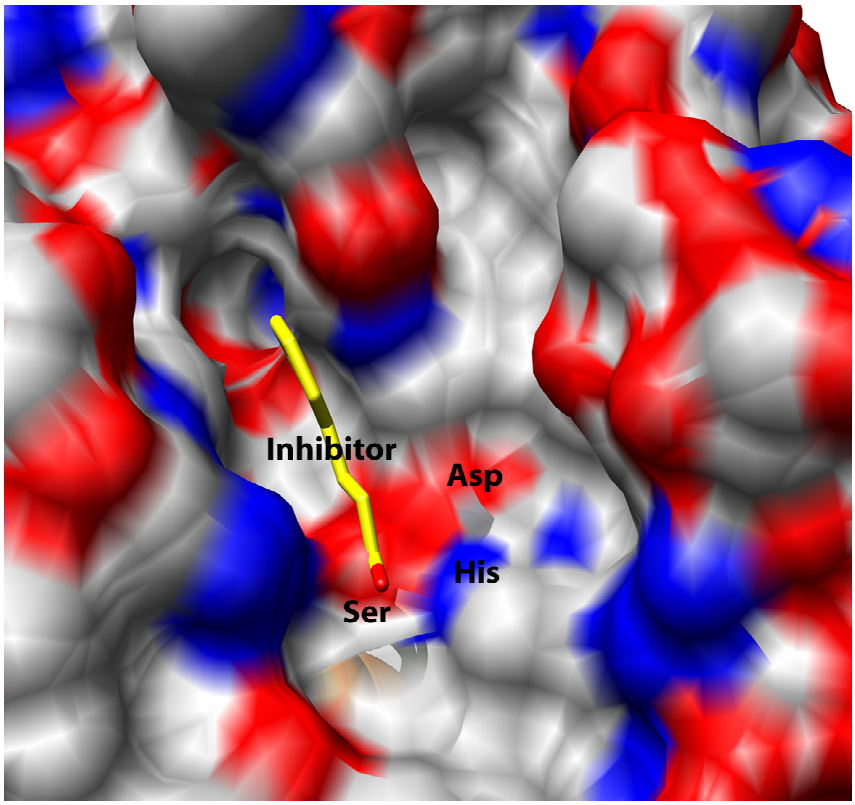


Fig 16. **Surface representation of the binding Pocket of mycP1 homology modeling.** The putative inhibitor binding packet of mycP1 docked with TCA (a serine protease inhibitor). Blue: Positively charged residues, Red: Negatively charged residues

Figure 16 shows a closer look at the binding pocket of the homology model. The first step for docking is to define the possible ligand binding sites (figure 16). The serine protease inhibitor TCA was docked into the binding cleft of the MycP1 model. The catalytic residues Asp90, His121 and Ser332 were used as the anchor residues for this docking. This predicted binding pocket could be used in the virtual docking by screening more than 2 million ZINC small molecule compounds from the MODELLER6.0 database.

# CHAPTER IV

# SUMMARY AND CONCLUSIONS

1. *Mycobacterium tuberculosis* contains all five mycosin proteins. However, to date, MycP1 is thought to be the most significant of the five because it is expressed during infection. The expression points of the other four mycosin proteins are not really known at this time. They are all subtilisin like serine proteases and have highly conserved active sites, as shown in figure 3. This enables the characterization of MycP1 to be applied to all five *Mtb* mycosins.

2. Drug design aims to propose an inhibitor that will enter and bind tightly to the target's active site. This will prevent the target's substrate from binding and therefore being processed. It is important to ensure that the target's affinity for the inhibitor is much greater than its affinity for the substrate. If the target does not bind the inhibitor tightly, it will possibly be dislodged temporarily, allowing the substrate to enter and be processed. Although, this processing would be greatly reduced, it would lower the effectiveness of the inhibitor as a potential treatment.

The essentiality of proteases to the survival of organisms makes them attractive drug targets. Proteases are a super family of enzymes that act by hydrolyzing peptide bonds in other proteins. They make up approximately 1-5% of an

organism's gene content. Proteases are essential to the normal functioning of cells because they perform a number of post-translational cleavages. Many proteins will not reach their active form if not processed by a protease. This will inevitably disrupt normal functioning and cause "junk" proteins to build up in the cell. Depending on the essentiality of the specific protease function, the cell will then die. Effectively inhibiting the function of an essential protease is a novel way of affecting several proteins at once. These unprocessed, non-functional proteins will disrupt the cell's life cycle at targeted places, effectively treating an infection.

3. The conservation of the active sites among the five mycosin proteins increases the interest in MycP1 as a drug target. The most potent inhibitor will, presumably, bind to all five mycosin proteins, inactivating them. This will enable a multi-positional attack on the infection. However, to test this, the mycP2 gene has also been cloned for use in expression. MycP2 will be used to test inhibitors along side MycP1 to see the actual activity.

4. The homology model generated for MycP1 can be used as a tentative receptor model in an *in silico* virtual screening using a small molecule database. The database will fit over two million possible inhibitors to the active site of the homology model. The top thousand hits from this virtual screening can then be used to test the actual activity of the protein. It is absolutely essential to test all

potential inhibitor with the purified mycosin protein. The homology model is only an estimation of the structure of MycP1. The actual structure will differ from this model, potentially changing the binding of an inhibitor drastically. Although no homology modeling of MycP2 will be done virtually, the active protein will be tested along side MycP1 to compare the inhibitor binding.

5. The expressed Mycp1 protein can be used in two ways:
    a. It is important to test the MycP1 protein with potential inhibitors using an activity assay. This will test the ability of MycP1 to cleave the peptide substrates in the presence of inhibitors. Para-nitro annelid incorporated short peptide fragments can be used as substrates. When MycP1 cleaves the peptide, the para-nitro annelid will be release into the solution. This free para-nitro annelid can be detected using a spectrophotometer. The most ideal inhibitor will prevent MycP1 from cleaving the peptide and therefore result in a low absorbance in the spectrophotometer.
    b. Crystallization of the MycP1 protein will allow the three dimensional structure to be solved. The MycP1 crystal will be shot with an x-ray beam, creating a diffraction pattern. Using computer analysis, the three dimensional structure will be solved. This will enable us to examine the actual active site of the MycP1 protein. Virtually, inhibitors can be docked to determine the best candidates.

The MycP1 protein can also be co-crystallized with potential inhibitors. Soaking MycP1 crystals in inhibitor solutions will allow the inhibitor to enter and bind to the protein. The soaked crystal can then be shot with an x-ray beam. The inhibitor will show up in the diffraction pattern, demonstrating the interaction with the active site of the protein. With this information, the inhibitor can be virtually modified to optimize binding.

# REFERENCES


1. Organization, W. H. (2005)
2. Van Rie, A., and Enarson, D. (2006) XDR tuberculosis: an indicator of public-health negligence. In. *Lancet*, 368 Ed.
3. Arnold, C. (2007) *Clinical Microbiology & Infection* **13**(2), 120-128
4. Reinhardt, E. (2005) *UN Chronicle* **42**(2), 17-19
5. Brown, G. D., Dave, J. A., Gey van Pittius, N. C., Stevens, L., Ehlers, M. R., and Beyers, A. D. (2000) *Gene* **254**(1-2), 147-155
6. Thompson J.D., H. D. G., Gibson T.J. . (1994.) *Nucleic Acids Res.* (22), 4673-4680
7. Beitz, E. (2000) *Bioinformatics* **16**(135-139)
8. Converse, S. E., and Cox, J. S. (2005) *Journal of bacteriology* **187**(4), 1238-1245
9. Server, v. T.
10. Altschul S.F., G. W., Miller W., Myers E.W., Litman, D.J. (1997) *Journal of molecular biology* **215**, 403-410
11. Siezen, R. J., and Leunissen, J. A. M. (1997) Subtilases: The superfamily of subtilisin-like serine proteases. In.
12. Scordis, P., Flower, D. R., and Attwood, T. K. (1999) *Bioinformatics* **15**(10), 799-806
13. Marchler-Bauer A, A. J., Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH. (2007) *Nucleic Acids Res.* **35**, D237-240
14. B Rost, G. Y., J Liu. (2004) *Nucleic Acids Research 32* **(Web Server issue)**(W321-W326)
15. Guex N, P. M. (1997) *Electrophoresis* **18**, 2714-2723
16. Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999) *Journal of molecular biology* **294**(4), 1027-1040

# CONTACT INFORMATION


| | |
|---|---|
| Name: | Hilary Jean Baird |
| Professional Address: | c/o Dr. James C. Sacchettini<br>Department of biochemistry<br>Texas A&M University<br>College Station, TX 77843 |
| Email Address: | hil.baird@gmail.com |
| Education: | BS Biochemistry, Texas A&M University, May 2007<br>Undergraduate Research Scholar |